1 Hand-written Part

P1

$$\varphi'(s) = \theta(s) + s\,\theta'(s)$$
$$= \theta(s) + s\cdot\frac{e^{-s}}{(1+e^{-s})^2}$$
$$= \theta(s) + s\cdot\theta(s)\times\frac{e^{-s}}{(1+e^{-s})}$$
$$= \theta(s) + s\cdot\theta(s)\cdot(1-\theta(s))$$
$$= -s\theta^2(s) + (s+1)\theta(s)$$

P2

(A)

$$V_0 = \left[\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right]^T$$
$$V_1 = PV_0 = \left[\tfrac{1}{2}, \tfrac{1}{6}, \tfrac{1}{3}\right]^T$$
$$V_2 = PV_1 = \left[\tfrac{1}{3}, \tfrac{1}{6}, \tfrac{1}{2}\right]^T$$
$$V_3 = PV_2 = \left[\tfrac{5}{12}, \tfrac{1}{4}, \tfrac{1}{3}\right]^T$$
$$V_4 = PV_3 = \left[\tfrac{5}{12}, \tfrac{1}{6}, \tfrac{5}{12}\right]^T$$
$$V_5 = PV_4 = \left[\tfrac{3}{8}, \tfrac{5}{24}, \tfrac{5}{12}\right]^T$$

(B) Let $V^* = [X, Y, Z]^T$

$$V^* = PV^* \Rightarrow \begin{cases} X = Y + \frac{1}{2}Z \\ Y = \frac{1}{2}Z \\ Z = X \end{cases}$$

$$2X = Y = 2Z$$

$$V^* = \left[\tfrac{2}{5}, \tfrac{1}{5}, \tfrac{2}{5}\right]^T$$

P3.

(A)

$dis((1,2),(7,0)) = \sqrt{40}$

$dis((3,4),(7,0)) = \sqrt{32}$

$dis((1,2),(10,2)) = \sqrt{81}$

$dis((3,4),(10,2)) = \sqrt{53}$

(1,2)	(3,4)
(1,2)	(3,4)
	(7,0)
	(10,2)
(1,2)	$(\frac{20}{3}, 2)$

(1,2)	$(\frac{20}{3}, 2)$
(1,2)	(7,0)
(3,4)	(10,2)
(2,3)	$(\frac{17}{2}, 1)$

(2,3)	$(\frac{17}{2}, 1)$
(1,2)	(7,0)
(3,4)	(10,2)
(2,3)	$(\frac{17}{2}, 1)$

converge

(B)

(1,2)	(7,0)
(1,2)	(7,0)
(3,4)	(10,2)
(2,3)	$(\frac{17}{2}, 1)$

(2,3)	$(\frac{17}{2}, 1)$
(1,2)	(7,0)
(3,4)	(10,2)
(2,3)	$(\frac{17}{2}, 1)$

converge

No, the result is same as (A)

(C)

(1,2)	(3,4)
(1,2)	(3,4)
	(5,6)
	(7,0)
	(10,2)
(1,2)	$(\frac{25}{4}, 3)$

(1,2)	$(\frac{25}{4}, 3)$
(1,2)	(5,6)
(3,4)	(7,0)
	(10,2)
(2,3)	$(\frac{22}{3}, \frac{8}{3})$

(2,3)	$(\frac{22}{3}, \frac{8}{3})$
(1,2)	(5,6)
(3,4)	(7,0)
	(10,2)
(2,3)	$(\frac{22}{3}, \frac{8}{3})$

converge

(1,2)	(10,2)
(1,2)	(7,0)
(3,4)	(10,2)
(5,6)	
(3,4)	$(\frac{17}{2}, 1)$

(3,4)	$(\frac{17}{2}, 1)$
(1,2)	(7,0)
(3,4)	(10,2)
(5,6)	
(3,4)	$(\frac{17}{2}, 1)$

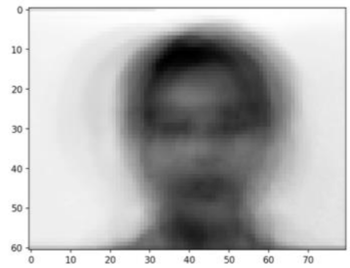
converge

(1,2) (3,4) converge to (2,3) $(\frac{22}{3}, \frac{8}{3})$ which is different from (1,2) (10,2) converge to (3,4) $(\frac{17}{2}, 1)$ *

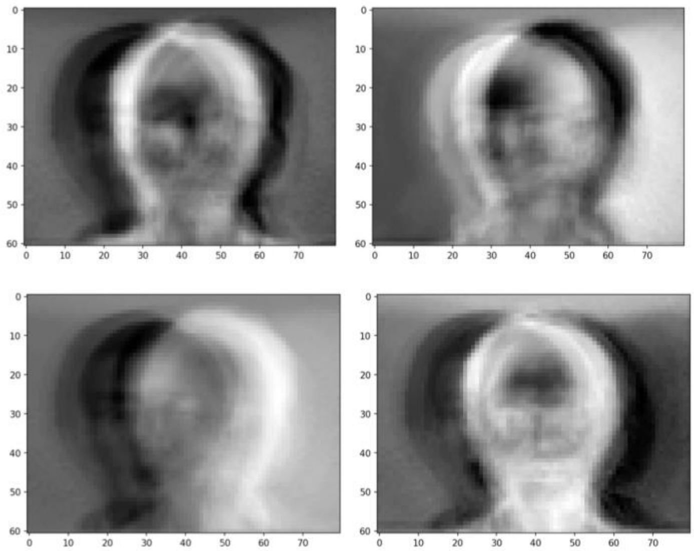
2 Programming Part

(a)

Mean

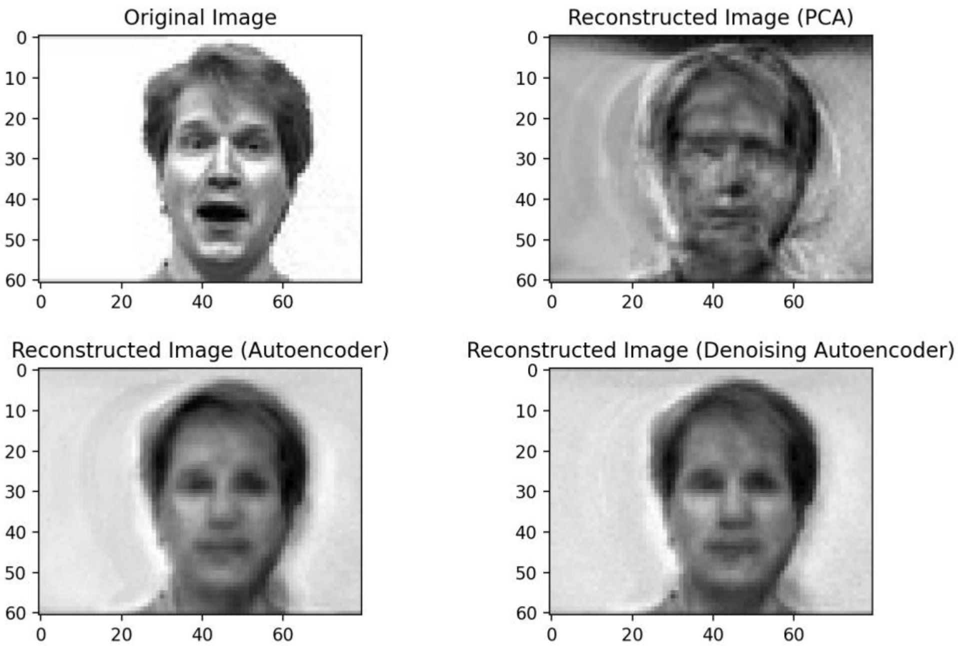


Top 4:

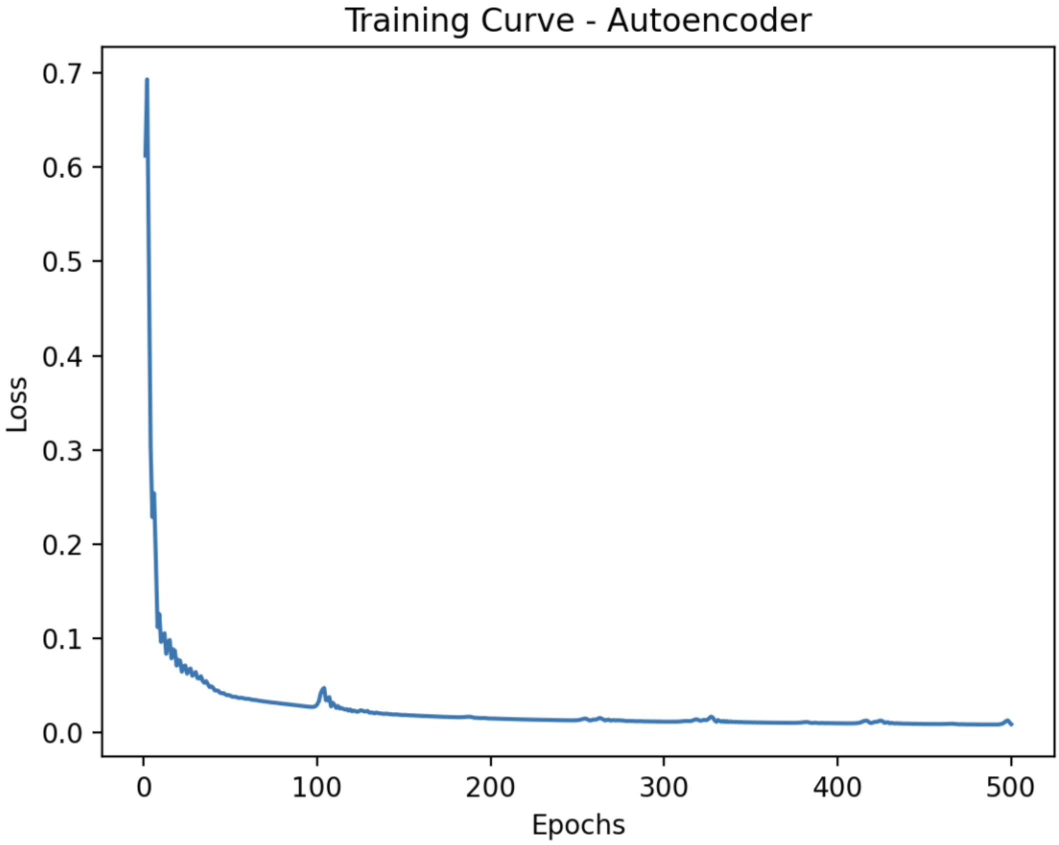


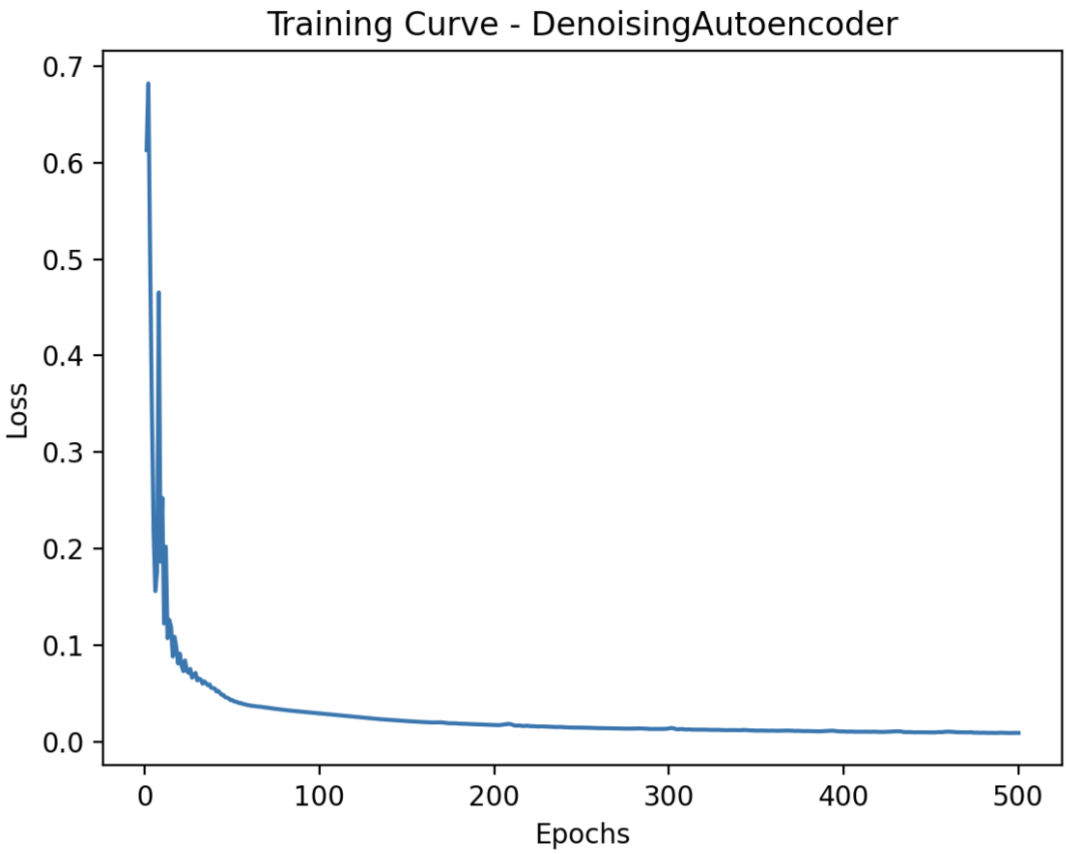
(b)

MSE: PCA = 0.0107, Autoencoder = 0.0148, Denoising Autoencoder = 0.0138



(c)


Training Curve - Autoencoder
Loss
Epochs
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
100
200
300
400
500


Training Curve - DenoisingAutoencoder
Loss
Epochs
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
100
200
300
400
500

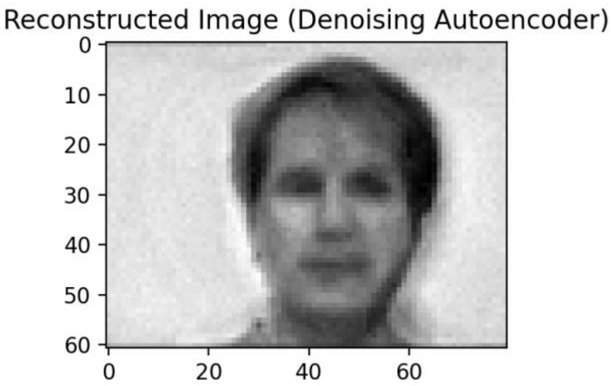
(d)

Shallower:

Encoder: Linear(4880, 488) -> Linear(488, 244) -> Linear(244, 122)

Decoder: Linear(122, 244) -> Linear(244, 488) -> Linear(488, 4880)

MSE: 0.0137

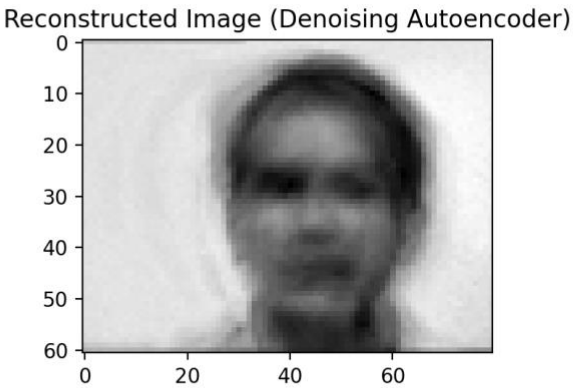


Deeper:

Encoder: Linear(4880, 1952) -> Linear(1952, 488) -> Linear(488, 244) -> Linear(244, 122)

Decoder: Linear(122, 244) -> Linear(244, 488) -> Linear(488, 1952) -> Linear(1952, 4880)

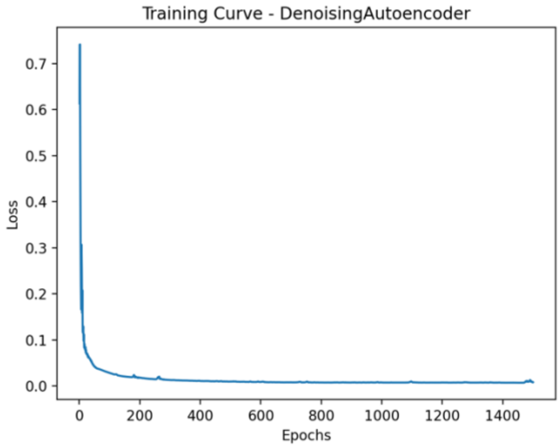
MSE: 0.0191

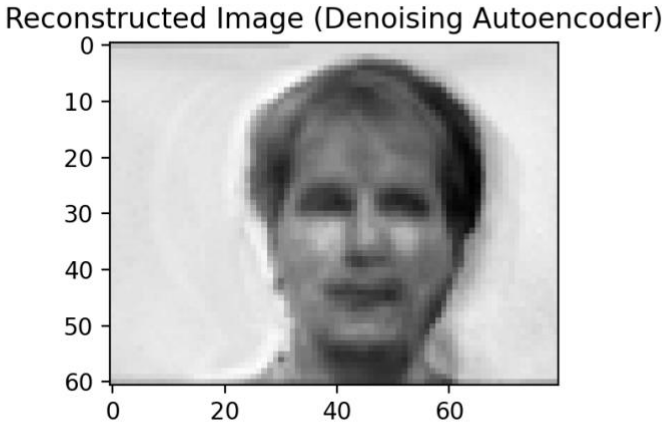


We can see that deeper model does not perform better in this case. The reconstruction error for shallower model is lower.

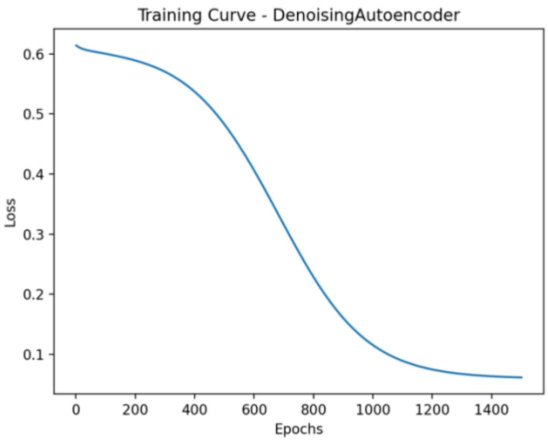
(e)

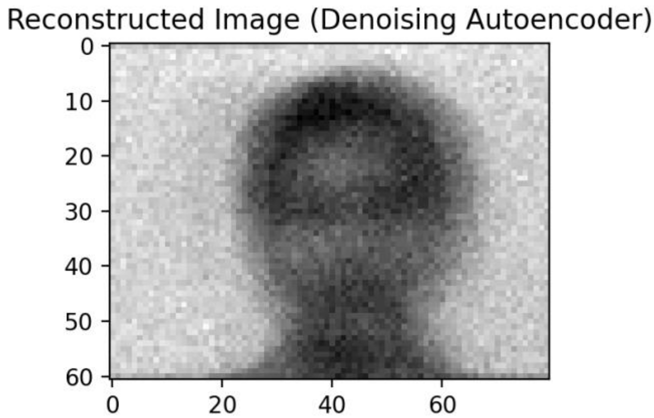
Adam:





SGD:





From the training curves above, we can see that Adam has faster convergence speed and is able to find better local minimum; the reconstructed image also tells us that Adam(mse = 0.0132) performs better than SGD(mse = 0.0374).